government viewpoint of clinical trials

by Robert Temple, M.D.

I. INTRODUCTION

Clinical trials are prospective, organized, systematic exposures of patients to an intervention of some kind (drug, surgical procedure, dietary change) in order to answer some question about the intervention. Clinical trials are not the only way to find things out; retrospective studies, alert but unplanned observation, and epidemiologic analyses all can teach us something and have their place, but the clinical trial is unique in being under the investigator's control, subject not to data availability or chance but to his ability to ask good questions and design means of answering them.

Clinical trials have both a scientific and regulatory role. Well-controlled clinical trials are, by law, a necessary part of the substantial evidence of effectiveness that must be adduced before a drug may be marketed. At the same time, it is the requirements of science, on which the law is based, that make clinical trials the principal means by which the beneficial and adverse effects of drugs are discovered and intelligent directions for use of drugs are developed. Clinical trials of drugs should not be thought of as burdens that must be borne only, or mainly, "to satisfy FDA." They are performed to find out the answers to pertinent scientific questions and in most cases FDA and drug scientists would reach substantial agreement on what those questions are. What, then, are the difficulties that arise in trying to identify pertinent questions and answer them through well-designed trials? The difficulties are, broadly, of two kinds:

1. Individual studies may be designed without careful attention to the questions they really are capable of answering. The result is either a) a useless trial that answers no question at all or, b) a trial that answers some other question, not the one intended, or only part of the intended question.

2. The total package of studies may be designed without a thoughtful consideration of all the questions that are pertinent. There are, of course, practical limitations on the number of studies that can reasonably be expected; nevertheless, it seems possible that more of the pertinent questions can be answered without any increase in the total number of patients exposed in clinical trials.

II. PROBLEMS IN DESIGN OF INDIVIDUAL STUDIES

The optimal study design depends on the question asked; the ability to carry out the optimal study may be limited by ethical, practical, and economic considerations. Still, a study must be sufficient to its task, and design limitations should be understood before proceeding, first to see whether a better design can be found and to understand the limits on interpretation imposed by a less than optimal design, and second, so that

Dr. Temple *is with the Bureau of Drugs, Food and Drug Administration, Rockville, MD.*

Presented at the Drug Information Association Workshop, December 9, 1981, Williamsburg, VA.

essary, the limits can be discussed with the regulatory ency and potential problems anticipated. Let me ustrate this with several specific examples of design oblems we have seen and of alternative designs that ould be considered, but rarely are.

Positive control studies

A well-controlled clinical trial, that is, one with dequate methods to reduce bias, such as double-blinding and randomization to groups, can use several kinds of ontrols. FDA regulations provide that instead of a placebo control "an effective regimen of therapy may be used for comparison, e.g., where the condition treated is such that no treatment, or administration of a placebo would be contrary to the interest of the patient."

I think no one would disagree with the idea that where treatment that is known to prevent death or serious morbidity exists, it is not ethical to deny patients the proven benefit of such treatment while studying a new therapy. (I will ignore as beyond the scope of this paper the very difficult question of whether it is ethical to study *any* new therapy in that case unless there is good reason to hope it will be superior to the established treatment.) Sometimes in such cases it may be possible to compare the new therapy with a placebo when they are added to the proven treatment, but in many cases this may not be appropriate, e.g., if the agents are pharmacologically similar, and the only recourse is to compare one therapy with the other.

This sounds like a reasonable and straightforward thing to do, yet it is fraught with theoretical and practical problems. I should emphasize that these problems arise only when the hoped-for outcome is a finding of no difference between the therapies. A positive control trial in which the objective is to show a difference, i.e., to *beat* the control, (this is typically the case in a dose-response trial, where the higher dose is expected to be superior to the lower dose) is conceptually and analytically similar to a placebo-controlled study and need not concern us further. Trials in which a placebo group *and* a positive control are used also do not pose problems and are an excellent design.

When the study objective is to show that a new drug is effective by showing its similarity to a positive control, several difficulties arise:

1. Lack of an agreed upon test for statistical significance.—For better or worse, the scientific community has generally agreed that a difference between two treatment groups will be considered statistically significant when the alpha-error, the chance of incorrectly rejecting the null hypothesis is 0.05 or less, i.e., there must be less than one chance in 20 that the apparent difference seen was a chance event. The FDA has accepted this standard and so has most of regulated industry. It is very convenient. There can be disagreements, of course, about 1-tail vs. 2-tail tests, about corrections for multiple comparisons, etc., but they usually are disagreements about whether the nominal finding of p less than 0.05 in a particular case is real; i.e., whether the accepted standard has truly been met.

There is no comparable convention for agreeing upon "statistically significant similarity." Here, of course, the crucial question for a regulatory agency is not: "What is the chance of incorrectly rejecting the null hypothesis of no difference?" as no difference has been shown, but "What is the chance of incorrectly failing to reject the null hypothesis when the drugs are really different," i.e., what is the beta-error? The size of the beta-error is not absolute but depends on the difference one would like to detect, i.e., the beta-error is usually very large with respect to detecting a very small difference, such as 1%. It is smaller as the difference to be detected grows.

Although for certain purposes, particularly bioavailability studies, we have begun informally to think about a 10-20% beta-error as acceptable if we are trying to detect a 20% difference; this really has not been sufficiently discussed. There is even less agreement when it comes to more complex clinical trials.

Difficult and important as it is, the beta-error question is not necessarily the most difficult problem posed by positive control trials because it is a recognized problem, often considered by designers of clinical trials, and because analyzing the beta error is not difficult. There is a long way to go, however, as the question is still more often ignored than faced. Moreover, the result of an analysis of beta error is likely to be unpleasant. An 80 or 90% chance of detecting a 20% difference, were that to become an accepted standard, will in most clinical situations require a very large study.

2. The second problem with positve-control studies is that the incentives are wrong—In a placebo-controlled trial, a large beta-error is definitely not in the interest of the sponsor of the trial if he is seeking to demonstrate the effectiveness of test treatment. The trialist has powerful incentives to minimize the beta-error by reducing needless variability resulting from imprecision of measurement, interfering concomitant treatments or other influences, and by assuring an adequate number of subjects. The incentives are thus in the direction of study excellence because sloppiness and error obscure differences, and a difference is what must be shown.

In a positive-control trial, however, *lack of difference* is what is sought. Sloppiness still obscures differences, of course, leading to a lack of an important incentive toward excellence. Although I cannot document the consequences of this in human studies, and perhaps the quality of investigators is such that in most cases there is no real cause for concern, past experience with chronic animal testing seems relevant and is not reassuring. In such animal studies, lack of difference from placebo is the "favorable" outcome, just as it is in the positive control clinical study. It is clear that in the past, even major pharmaceutical houses tolerated very poor studies, sometimes in their own laboratories and more often in the laboratories of their contractors. I have no reason to think this occurred by design but it seems possible at least one of the reasons for it might have been the absence of a strong push toward excellence.

The incentives problem is in part related to the beta-error question. If we are vigorous in asking how statistically meaningful a finding of lack of difference is, it will become apparent that small n's and needless variation

both obscure findings and lead to a large beta-error.

3. The third difficulty, and one that has never been discussed adequately, is that there is an unstated, often unrecognized, assumption underlying the use of a positive control that is not necessarily correct and cannot be assumed without careful analysis.—Showing that two drugs are equivalent in a study does not demonstrate that either is effective; it shows that both were effective or that neither were. Because the positive control is known to be an effective agent, we usually *conclude,* if equivalence is shown, that both agents were effective. This seems reasonable at first glance, but a closer look reveals a crucial assumption underlying that conclusion, viz., that the effective drug was effective *in the particular study in question.*

Recognizing this assumption, we can consider whether it is valid. In fact, it is not necessarily valid, because many effective agents are not demonstrably effective every time they are tested. It is not uncommon, for example, for effective mild analgesics, anti-anxiety agents, anti-depressants, or appetite suppressants not to "beat the placebo" in particular studies, and this can also occur in studies of mild antihypertensives or antianginal agents. The failure of an effective agent to appear effective in a particular study can occur for a variety of reasons, including bad luck, a high degree of placebo-responsiveness or spontaneous improvement in the study population, great variability of the study measurements, too small an n, or improper diagnostic criteria so that some of the subjects lack a condition responsive to the test drug. On the other hand, we would expect potent antihypertensives like guanethidine or effective antibiotics used against sensitive bacterial strains to be effective virtually every time they were compared with placebo.

If we cannot be very certain that the positive control in a study would have beaten a placebo group, had one been present, the fundamental assumption of the positive-control study cannot be made and that design must be considered inappropriate under the circumstances.

When I first considered this question in 1978 I was unable to find any explicit discussion of the issue, although it was clear that experts in certain areas recognized the principles involved. Analgesics, for example, are almost invariably studied using a placebo or dose-response methodology. I doubt any modern analgesiologist would try to show the effectiveness of a new analgesic by comparing it with aspirin in a single dosage-level study without a placebo.

It may not be surprising, then, that it was a pain specialist who recently described the difficulty in interpretation posed by positive control trials in a wonderfully terse and lucid manner. Writing in the *European Journal of Clinical Pharmacology*, Lasagna, after noting that in life-threatening situations, where known effective treatment exists, placebos cannot be used and the allocation of patients even to an unproven remedy must give us pause, said:

> In serious but less critical medical situations, one can justify a comparison between new drug and standard, even if a placebo group seems out of the question. But such a trial is convincing only when the new remedy is superior to *standard* treatment. If it is inferior, or even indistinguishable from a standard remedy, the results are not readily interpretable. In the absence of placebo controls, one does not know if the "inferior" new medicine has any efficacy at all, and "equivalent" performance may reflect simply a patient population that cannot distinguish between two active treatments that differ considerably from each other, or between active drug and placebo. Certain clinical conditions, such as serious depressive states, are notoriously difficult to evaluate because of the delay in drug effects and the high rate of spontaneous improvement, and even known remedies are not readily distinguished from placebo in controlled trials. How much solace can one derive from a trial that shows no difference between a new putative antidepressant and a standard tricyclic?

It is clear from experiences in the past that drug sponsors and investigators usually do not consider this problem at all.

Some years ago, the FDA was involved in controversy over the appetite suppressant agents, amphetamines and their relatives. After review of hundreds of studies we concluded the drugs did contribute to weight loss to a small degree, a fraction of a pound per week beyond what was achieved by placebo. Given this limited effectiveness, it is not at all difficult to conduct a study in which the test drug is not superior to placebo, and many such studies are in our files and in the medical literature. It might seem self-evident that it would not be reasonable to try to demonstrate the effectiveness of a new anorectic agent by comparing it with, for example, dextroamphetamine, but FDA spent many months on a case that turned on that very point.

Dexamyl was a combination of dextroamphetamine and amobarbital, the barbiturate intended to minimize the amphetamine side effects without, of course, eliminating the anorectic effect. To show that the anorectic effect was unaltered dextroamphetamine and Dexamyl were compared in a multicenter parallel comparative study without a placebo group. The weekly weight loss on both drugs was similar, but was within the range of weight loss commonly seen in placebo groups in weight loss studies. FDA argued that a showing of similar weight loss in the two groups was without meaning and could not show the effectiveness of either agent. The drug's sponsor failed to consider this question at all beyond asserting that dextroamphetamine is considered effective by FDA so that showing Dexamyl is equivalent to it must consitute proof of effectiveness.

A few years ago FDA refused to approve a new beta-blocker for use in angina because all of the controlled trials used a positive control design. Both the test drug and the control (propranolol) produced a reduction in angina rates and increased treadmill exercise compared to baseline but we were not able to conclude *a priori* that the response seen was beyond what might have been seen in a placebo group. This is a closer question than the analgesic or anorectic cases because the great majority of placebo-controlled trials (at least of published trials) of beta-blockers in angina do show effectiveness.

In such a case it may be reasonable to use a positive control, but this cannot simply be presumed.

Anyone proposing to demonstrate effectiveness by demonstrating equivalence to a known effective agent

must assume the burden of showing that the fundamental assumption of the positive control study is correct. I have no illustration of what such a showing should consist of, because so far as I know no drug firm has ever attempted to provide one, but I believe it would include the following:

A. A review of known placebo-controlled studies of the proposed positive control to show that the drug regularly can be shown superior to a placebo.

B. Close attention to the design and conduct of the studies in which the proposed positive control was shown to be effective so that the proposed new study can be designed similarly, utilizing a similar patient population and similar procedures (such as baseline procedures, dose, dosage regimens, titration method, training procedures, response assessment methods, and control of concomitant therapy).

C. An estimate of the size of the placebo response in studies of design similar to the proposed design and specification of a level of response in the treatment groups that would be considered clearly beyond what could be attributable to placebo.

There is no doubt that the reluctance to expose patients to placebo periods of any substantial length will not diminish and that the pressure to utilize positve-control studies to demonstrate effectiveness will grow. If we are not to suffer a serious degradation in the quality of evidence used to support applications the fundamental assumption underlying the positive control design must be recognized and its validity in any particular case addressed. It appears likely that there will be some drugs for which a positive control design will never be appropriate (e.g., minor tranquilizers, anorectic agents, or antidepressants), others where it is appropriate and almost mandatory (antibiotics in serious illnesses), and a large middle group where this design may be appropriate if properly supported.

None of the above, of course, means that there is anything wrong with a positive-control study. When these are used to answer the right questions, they are the right studies. In fact, with few exceptions, I believe every new agent should be compared to similar agents in studies large enough to determine with some precision whether there are important differences in side effects and effectiveness. In any such study, however, it is extremely helpful to include a placebo treatment group (even if it is maintained only briefly and even if it is smaller than the other groups) to confirm the validity of the assay in the population studied. This rather obvious design is too rarely employed.

B. Titration

It is understandable that many clinical trials would include a period of titration. This is a procedure that conforms to good clinical practice and seems to satisfy the demand that patients be exposed only to the amount of drug they need. Moreover, it seems efficient, because it gives the drug its best chance by assuring use of an adequate dose and because it seems able to provide dose-response information. Unfortunately, the usual methods of titration commonly lead to confounding of drug effects with time effects. Where titration is used to be sure that *at least* an adequate dose is used, e.g., in the early trials to be certain the drug is effective, the procedure does not interfere with interpretation. Where it is used to define dose-response relationships or to compare potency of two drugs, however, it can lead to fallacious results and tends to overestimate the amount of drug needed.

I believe this has been happening consistently in the area of hypertension and it would not surprise me if it were the case elsewhere. We recently reviewed data on the cardio-selective beta-blocker atenolol. Early large studies in which patients were titrated to some endpoint, then maintained, used doses of 600 mg per day and more. Subsequent controlled parallel comparisons of several doses have shown that doses larger than 100 mg per day have no additional effect and that it is not easy to distinguish 50 mg from larger doses.

The history of diuretic dosage in hypertension is similar, at least in one well-defined case. Chlorthalidone for years has been used in a dose of 100 mg per day. Two recent well-controlled studies[2-3] have shown that 25 mg is a fully effective dose.

Hypertension is perhaps especially liable to the distortions introduced by titration procedures because patients selected for study regularly tend to have a spontaneous fall in blood pressure during a trial. Certain drugs, too, such as beta-blockers or diuretics, may be especially susceptible to excessive dosage as they have relatively little dose-related toxicity (once beta-blockade is achieved or maximum sodium excretion attained). There is thus not much incentive to minimize the dose.

The tendency of titration to overestimate the necessary dose can be illustrated by recent experience with captopril, and in this case the high doses used in most patients may have led to dose-related toxicity. The mean diastolic blood pressures for the major captopril-placebo comparison are shown below:

	Week				
	0	**1**	**2**	**3**	**4**
Captopril	110	100	99	96	94
Placebo	110	104	104	103	101
C/P Difference	0	4	5	7	6

According to the titration rules, the dose was increased weekly from 25 mg t.i.d. up to 150 mg t.i.d until a diastolic pressure of 90 mmHg was reached. About half of the patients reached the 150 mg t.i.d. dose and about 70% had at least 100 mg t.i.d.

Notice, first, that the placebo group did quite well in this study and, second, that the drug-placebo difference does not change much from week 2 to 4. But the actual diastolic pressure fell nicely from week 2 to 4, giving the titrating clinicians the impression that they were getting an increased response. As a result of this and other studies the impression was gained that larger doses were more effective. In fact, small doses, 25 mg t.i.d., block the angiotensin-converting enzyme quite well and, perhaps

with some exceptions, 50 mg t.i.d. seems to treat most hypertensive patients, even those with severe, resistant hypertension.

The perception that response increased with doses up through 450 mg/day led to use of these doses in resistant patients, including those with complicated hypertension, many of whom had renal disease. Unfortunately, in those patients there was a comparatively high rate of agranulocytosis, about 2-3% (depending on which patients are included in the denominator). Whether this occurred because the drug somehow interacted with the underlying diseases of these patients (lupus, etc.) or reflected the high dose used compounded by further blood level elevation due to renal impairment[4] cannot be said with certainty, but the latter seems, to me at least, more likely. Perhaps consistent use of lower doses and modification of dose for renal impairment could have averted these problems. Some drugs are unlikely to be given in excessive doses because dose-related toxicity is an obvious problem (e.g., disopyramide) or because the drug is regularly titrated to the point of intolerance (e.g., anticholinergic drugs or adrenergic neuron blocking drugs). Where this is not the case, the impressions gained from sequential titration studies should be examined in a formal dose-response study in which patients are randomly assigned to several doses, e.g., the dose thought to be the usual effective dose, a multiple of that dose and one or more fractions of the dose. This should be done early in the course of drug development.

C. Failure to control for change in a critical baseline characteristic and failure to examine such characteristics in properly designed trials.

A common study design involves selection of a patient group with some important baseline characteristic, e.g., failure to respond to a prior therapy or intolerance to prior therapy, and then switching the patients to a new treatment, often a drug of the same class. Typically, patients respond to the new treatment or tolerate it better, leading to a claim for better tolerance than the previous drug or "usefulness in resistant cases."

Interest in patients with these characteristics is highly appropriate, of course, but the study design is badly flawed unless there is assurance that the adverse effect would have persisted had therapy not been changed or that effectiveness of treatment could not be influenced by such things as improved compliance, better attention to ancillary therapy (e.g., salt intake in patients with heart failure) or passage of time. A proper study of such patients would involve *selecting* the patients because of their poor response or an adverse effect on the "old" drug, then randomly assigning them to the new agent and the old one. This cannot be done, obviously, if the adverse effect is very dangerous.

I should note that even this design, while better than the usual study, has a bias and must be narrowly interpreted. The design is biased against the first drug as, by definition, all patients on it have done badly. The second drug is unlikely to do worse, or even as poorly. This kind of study thus should not be construed as a comparison of the effectiveness or adverse effects of two drugs. For example, if drug B caused half as many nightmares as drug A in patients selected because of having nightmares on drug A, it is still possible that drug A would also cause half as many as drug B in patients selected because of having nightmares on drug B. The study can, however, determine whether there is any reason to try drug B in patients doing badly on drug A and can tell you what to expect when you do so. It also can give a reason to proceed with the much harder and larger study comparing the drugs with each other in an unselected population.

There are numerous important questions that can, and should, be answered by studies like these. Do, for example, some beta-blockers have less ability to cause nightmares, depression, fatigue, cold extremities, claudication, or asthma? The pharmacology of the drugs suggests there *might* be such differences, but there are few studies of proper design to test the question. Why not? Well, it is easier perhaps to think optimistically about the possibly valuable pharmacologic differences between these agents than to study the actual clinical consequences of such differences. Not only that, the side effects are often evanescent and hard to study. One sponsor for example, planning to see whether his beta-blocker had less propensity to cause unpleasant dreaming than propranolol, found that only about 10% of patients with that complaint still had it when brought in for study. This illustrates the difficulty of these studies but it also shows how meaningless a study would be that simply switched patients to the alternative drug. A 90% "success" rate would be achieved even if the drugs were identical.

In one recent case, use of the correct design to compare therapies was extremely important. In evaluating effectiveness of captopril in populations resistant to other therapy, it would have been tempting to take the poor responders and simply switch them to captopril, the study design that was in fact initially suggested. Ethical concerns encouraged this, as the patients had unacceptable levels of blood pressure on previous therapy and were in fairly urgent need of treatment. At our urging, however, the sponsor developed a study design that minimized patient exposure to unacceptably high blood pressures, yet confirmed the failure to respond to previous therapy while showing effectiveness of captopril. In brief, failures on standard triple therapy (diuretic, beta-blocker, and hydralazine, a potent and generally well-tolerated regimen for resistant hypertension) were observed briefly on triple therapy and then randomized to the triple therapy or to captoril if blood pressure exceeded certain limits. There were many "escape clauses" for unacceptably high pressures, with resulting shifting populations, but in the end there was a valid comparison of triple therapy vs. captopril-containing regimens (captopril, usually with a diuretic, sometimes with a beta-blocker) that showed a clear advantage for the latter, with diastolic pressures of less than 90 mmHg attained in 46% of captopril patients vs. 14% of triple therapy patients, and a fall of at least 10% in diastolic pressure attained in 72% of captopril patients vs. 24% of triple therapy patients. Note, however, that one fourth of patients apparently unresponsive to triple therapy responded to it under the conditions of the trial and 15% were fully controlled.

The study designs described are all variations

me. In each case, the entry population is "enriched" h cases of particular interest to make a relatively common patient characteristic (an adverse effect or or response to vigorous alternative therapy) easier to dy. It would be very hard to compare directly the quency of two unusual adverse effects (e.g., night-res or intermittent claudication) with two beta-block- because you would need hundreds of patients per up to elicit the small fraction that have such effects detect a difference between two agents.

A similar technique could perhaps be used to identify opulations for study where response is hard to assess or ere only a small portion of patients may respond. It s been very difficult, for example, to demonostrate fectiveness of any antispasmodic agent in treating nctional bowel disease, perhaps because the disease is ultietiologic and has diverse and fluctuating manifesta-ons. Nonetheless, many highly qualified clinicians think umber of drugs are effective in these conditions. There possible reasons for this. Perhaps the good re-onders do not get into studies at all; they might just fill the prescription that worked the last time and not be en again, leaving for studies only more recalcitrant tients. If I were a drug company faced with a need to ow effectiveness for such drugs I would ask the best GI nicians I could find to locate patients who, in their dgment, respond to antispasmodics. These patients uld then be the subjects for formal, placebo-controlled ials.

2. Trial designs stimulated by ethical considerations.

The placebo-controlled trial is on the defensive on hical grounds, but the idea that such trials are not ways suitable is not a new one or even debatable. Where fesaving or life extending treatment is available pla-ebos cannot be utilized. More and more, however, we ee sponsors and investigators reluctant to use placebo-controlled designs whenever life-threatening illness is involved, whether or not there is good evidence that the test treatment or any other treatment is useful. We thus ee positive controlled studies where no agent is clearly effective or uncontrolled studies.

I would not minimize the importance of such ethical concerns, but there is also a societal cost when studies fail to answer important therapeutic questions definitively. If they recognize this cost investigators and manufacturers may find they can carry out a better study than at first seems possible. I have described above the captopril study in resistant hypertension, clearly a better study than the one first proposed. In a similar case, a randomized, placebo-controlled trial of a calcium antagonist was carried out in patients with unstable angina, although the trial initially proposed on eithical grounds would have simply given all patients the active drug. In this case the design used allowed patients to be rapidly crossed over to alternative therapy if they worsened or failed to improve during the first day of treatment. Such "early escape" designs, with inevitably unbalanced cross-overs (of course, the fact that they are unbalanced is evidence of drug effect) may be harder to analyze than a classic parallel or complete crossover study; but they are far more rigorous than a simple uncontrolled trial.

Another design, identified years ago as a "more ethical" kind of angina study[5] and encouraged recently by FDA in situations where conventional designs are difficult or impossible, is a randomized, placebo-controlled therapy withdrawal. In such a study only patients appearing to respond to treatment are exposed to the drug for any length of time, and at the end of a period of treatment patients are randomized to placebo or continued therapy. The substitution could be at the end of the study or at some randomly chosen intermediate time (a placebo "pulse"), perhaps an advantage where some change in a pharmacologic response needs to be hidden as well as possible. This design permits an early "escape" from ineffective therapy; i.e., as soon as some defined degree of deterioration is reached patients can be identified as failures and restored to active therapy. The failure frequency may be the principal effectiveness endpoint. This kind of design is especially helpful in studies of long-term effectiveness, where it may be difficult otherwise to determine whether a drug is having persistent effect without a concurrent long-term placebo treatment, which is usually unacceptable. In designs of this sort the possibility of rebound phenomenon must be considered, of course.

Finally, let me offer a controversial perception, namely, that crossover studies may have a more "ethical" feeling to them than parallel studies because they seem to represent a kind of "therapeutic trial." I realize that in both kinds of studies groups, rather than individuals, are usually analyzed and yet the crossover still does define to some degree the usefullness of a drug in a specific patient, knowledge that could very well benefit the particular patient, help alter or refine his therapy, etc. I would cite as evidence that my perception is shared, perhaps subliminally, by the common practice of telling patients in placebo-controlled trials that they can try the "real" drug afterward, in the long-term extensions that follow controlled trials. That is just an uncontrolled crossover, after all.

III. OVERALL CLINICAL PROGRAM

Individual studies of excellent design must still be part of a program that asks the "right" questions about a drug. Rightness, here, is necessarily an evolving concept, subject to changes not controlled by drug companies or regulatory agencies, but rather subject to the activities of the scientific community. It would be difficult to imagine approving a new anti-arrhythmic agent without some studies of its effectiveness in preventing programmed stimulation of arrhythmias, a concept in its infancy when the last oral antiarrhythmic drug approved for use in the US was evaluated.

Nonetheless, there are some kinds of questions that seem part of the evaluation of any drug, yet are sometimes ignored or studied inadequately. And only sometimes does one gain the impression that the evaluation of a drug, the package that comes to us in the form of an NDA is the result of a carefully considered master plan. I recognize that I am being unfair to some extent. First of all, we at FDA may see many applications for similar drugs and can from these pick out what is omitted from

each. It would be much more difficult for a single sponsor to anticipate everything that could be done. I realize also that a more perfect drug work-up might cost even more than present ones do, already a lot of money, but I am not sure this is the case. At present the numbers of patients included in an NDA is not determined by the numbers needed for studies to evaluate effectiveness but by the need to expose a certain number of patients (usually in the neighborhood of 750-1000) to the drug. It is possible, I think, to keep the exposure no larger, yet learn more. Finally, I think the most important reason NDAs seem unplanned may be that they are planned too early, before the implications of clinical pharmacology and early effectiveness studies can be digested and incorporated into designs for later trials. There is no simple solution to this except to encourage an attitude of flexibility throughout a drug's evaluation.

FDA has clinical quidelines for many drug classes and I would not want to restate them here. I would like, however, to mention several areas of investigation that are, in my experience, chronically underemphasized.

A. Dose-response

As a indicated earlier, dose-response information is frequently derived from titration studies or small numbers of patients in early open trials. The dose-response curve of a drug deserves formal controlled evaluation, unless the drug has a mode of administration that makes this unnecessary, e.g., if it is always titrated to the maximum tolerated dose because the maximum possible effect is always sought. (Anticholinergic agents have traditionally been given this way, probably because effectiveness endpoints have been very hard to measure.) In most cases physicians need to know the maximum useful dose, i.e., the dose beyond which there is little likelihood of further response. A parallel study comparing several doses with placebo can define the dose-response relationship. This need not, I should add, increase the total number of patients included in the NDA. The study can be an integral part of the overall safety evaluation, and is more informative than the kinds of multicenter studies now conducted, with 50-100 or more patients on placebo and on various titrated doses of drug.

There are other aspects of the dose-response relationship that are of interest. If possible, the relationship of response to steady state blood level should be examined to see whether blood level predicts response. The response should of course be measured on the rising part of the dose-response curve. The often-stated lack of correlation between beta-blocker blood levels and hypotensive response may well reflect failure to recognize that the relationship was examined at doses far in excess of the beta-blocking dose, i.e., on the flat part of every individual's dose-response curve.

B. Dose-interval

In the rush to increase the dose-internal in the treatment of many diseases, especially hypertension, where less frequent dosing is felt to enhance compliance, intelligent study design has often been ignored. It may be hard to believe, but many studies intended specifically to assess less frequent dosing intervals (even once daily dosing)have measured blood pressure a few hours after dosing, principally because that time suits clinic visit schedules, rather than just before dosing. Very few studies, even today, measure response at both peak and trough blood levels; perhaps this is a burdensome question to explore in an outpatient trial, as it requires that patients remain at the clinic for several hours, but it is surely pertinent.

Does less frequent dosing, which necessarily leads to a greater peak-trough difference, increase dose related side effects? A cardioselective beta-blocker, for example, with a half-life of 6 hours given once daily, will have a peak blood level about twice what it would be if the same dose were given in divided b.i.d. doses. Could this affect use of such a drug in asthmatics? As cardio-selectivity is not complete, could the increased peak blood level produce untoward bronchial effects? The question has not been explored, to my knowledge.

Tacked onto the tail end of a drug development plan, these questions can be difficult to answer. Assessed early, through careful open studies and by having one of the typical multicenter trials compare different dose intervals, it is not difficult at all, and, again, all of the data are pertinent to the overall safety information that needs to be gathered anyway. But dose-interval questions are sometimes the last thing considered. We have had the experience of reviewing, and meeting with a drug manufacturer about, a phase III plan for an antihypertensive agent that consisted entirely of studies with t.i.d. dosing regimens. Virtually as the meeting participants were walking out the door, we asked what dose interval would be recommended in labeling and were told twice daily. The manufacturer had intended somehow to link the clinical trials (t.i.d. dosing) and planned labeling (b.i.d. dosing) through a pharmacokinetic study, without having good blood level-response data or any trial with the proposed regimen.

C. Drug-drug and drug-disease interactions

Interactions are a sensitive matter because it can be said truthfully that there is almost no limit to the questions that could be asked. Nonetheless, it does not seem reasonable to market an agent without exploring by *in vitro*, animal or human studies, as appropriate, possible interactions with digoxin, anticoagulants, drugs that are metabolized by the liver, or drugs likely to be used concomitantly. Similarly every agent should be labeled to explain what dosage adjustments are needed for various degrees of renal impairment or whether the agent is dialyzable.

These questions are sometimes seen as additional requirements for manufacturers, added to an already lengthy development process, but considered early it might be very easy to answer them, sometimes as part, sometimes as spin-offs of studies already being conducted. In phase III trials, for example, it would be an easy matter to look at blood levels of a drug the patient was already receiving before and after therapy with a new agent. Even a relatively casual study could probably detect major changes, and minor alterations are probably not of interest.

Virtually any antihypertensive trial has patients

ond only partially to a new therapy. These patients ld be the subjects of a secondary study to explore the raction with other agents by randomizing them to ed placebo or added additional agent.

IV. CONCLUSIONS

least in the cardiovascular drug area, the require- is for a proper clinical trial are generally well- erstood by investigators and manufacturers. My meration of problems should not cause us to ignore enormous improvement in study design that has de the 1970's a revolutionary period with respect to study quality. Nonetheless, problems remain and are reasons to fear that progress could be undercut thical concerns, as we continue to develop effective apies for more and more diseases. We, therefore, t consider how good trial design can best be enhanced preserved. I have cited several problem areas and posed design solutions and suggested a number of ects of drug evaluation that could be improved with le or no increase in development time or cost.

REFERENCES

1. Lasagna, L.: Editorial: Placebos and controlled trials under attack. *Eur. J. Clin. Pharmacol.* 15:373-374, 1979.

2. Tweeddale, M.G., Ogilvie, R.I., and Rudey, J.: Antihypertensive and biochemical effects of chlorthalidone. *Clin. Pharmacol. Ther.* 22:519-527, 1977.

3. Materson, B.J., Oster, J.R., Michael, U.F., Bolton, S.M., Burton, Z.C., Stambaugh, J.E., and Morledge, J.: Dose response to chlorthalidone in patients with mild hypertension. *Clin. Pharmacol. Ther.* 24:192-198, 1978.

4. Onoyama, K., Hirakata, H., Iseki, K., Fujimi, S., Omae, T., Kobayashi, M. and Kawahara, Y.: Blood concentration and urinary excretion of captopril (SQ 14,225) in patients with chronic renal failure. *Hypertension* 3:456-459, 1981.

5. Amery, W. and Dony, J.: A clinical trial design avoiding undue placebo treatment. *J. Clin. Pharmacol.* 15:674-679, 1975.